

REMARKS

In the Office Action, the Examiner rejected Claims 1-28, which were all of the then pending claims, over the prior art, primarily U.S. Patent 6,405,161 (Goldsmith). Specifically, Claims 1-11, 13-15 and 21-23 were rejected under 35 U.S.C. 102 as being fully anticipated by Goldsmith; and Claims 12, 16 and 24 were rejected under 35 U.S.C. 103 as being unpatentable over Goldsmith in view of U.S. Patent 6,904,402 (Wang, et al.). Claims 17 and 25 were rejected under 35 U.S.C. 103 as being unpatentable over Goldsmith in view of U.S. Patent 7,103,536 (Kanno); and Claims 18-20 and 26-28 were rejected under 35 U.S.C. 103 as being unpatentable over Goldsmith in view of Kanno and Wang, et al.

Independent Claims 1, 13 and 21 are being amended to better define the subject matters of these claims. Claims 16, 17, 18, 24, 25 and 26 are being cancelled to reduce the number of issues in this case. Claims 19 and 20 are being amended to be dependent from Claim 13 instead of the now cancelled Claim 18; and, similarly, Claim 27 is being amended to be dependent from Claim 21 instead of the now cancelled Claim 25. Also, new Claim 29, which is dependent from Claim 13, is being added to describe a preferred feature of the invention.

For the reasons discussed below, Claims 1-15, 19-23 and 27-29 patentably distinguish over the prior art and are allowable. The Examiner is thus respectfully asked to reconsider and to withdraw the rejection of Claims 1-11, 13-15 and 21-23 under 35 U.S.C. 102 and the rejections of Claims 12, 19, 20, 27 and 28 under 35 U.S.C. 103, and to allow Claims 1-15, 19-23 and 27-29.

Generally, Claims 1-15, 19-23 and 27-29 patentably distinguish over the prior art because the prior art does not show or suggest refining sets of potential prefixes and suffixes, using knowledge of previously identified prefixes and suffixes, to find actual prefixes and suffixes, as described in independent Claims 1, 13 and 21.

The present invention, generally, relates to automatically collecting affixes of a language from one or more documents. As discussed in detail in the present application, knowledge of affixes is important for analyzing existing words and for producing new words. It is very difficult and time extensive to acquire a complete list of affixes of a language by hand, and various procedures have been tried to use a computer or some other automated process to identify affixes.

The previous approaches tend to parse words into pieces, either a prefix and a stem, or a stem and a suffix. Also, the prior art approaches tend to limit the length of an affix to reduce the size of the search space.

The prior art systems have a number of disadvantages and limitations. For instance, they may fail to discover both prefixes and suffixes at the same time, and they may not be able to discover nested affixes. Also, because of the limitation on length, the prior art systems do not find many affixes that appear in technical documents. Further, many of the previous approaches fail to find affixes containing non-alphabet characters such as digits and hyphens.

The present invention effectively addresses these prior art limitations. Generally, the present invention provides an unsupervised, knowledge-free procedure for automatically discovering prefixes and suffixes from text. The present invention integrates prefix and suffix discovery in such a way that uses knowledge about prefixes to find suffixes and uses knowledge about suffixes to find prefixes.

More specifically, in one embodiment, the invention provides a computer system for analyzing text in one or more documents, and comprising one or more system interfaces, and an affix process that determines one or more affixes of one or more words in one or more of the documents and provides the affixes to the system interface. This affix determining process comprises obtaining a collection of words, adding the words into a prefix Patricia tree, using said

prefix tree to identify a set of candidate prefixes, reversing each of the words, adding the reversed words into a suffix Patricia tree, and using the suffix tree to identify a set of candidate suffixes. The affix determining process further comprises refining the sets of candidate prefixes and suffixes to identify actual prefixes and suffixes, including using knowledge, previously discovered in the refining step, to further refine said sets of candidate prefixes and suffixes.

The references of record do not disclose or suggest refining sets of candidate prefixes and affixes as described above.

For instance, Goldsmith, the primary reference relied on by the Examiner, discloses an automated, morphological analysis of a natural language for determining prefixes, suffixes and stems. This analysis has three major components: determining the correct morphological split for individual words; establishing accurate categories of stems based on the range of suffixes that they accept and identifying allomorphs of the same stem.

There are a number of important features of the preferred embodiment of this invention that are not shown in or suggested by Goldsmith. For example, Goldsmith does not disclose or suggest the user of reverse words for identifying suffixes, or the use of Patrician trees to identify candidate sets of prefixes and suffixes.

Wang, et al, as the examiner observed in the Office Action, does disclose the use of a Patrician tree; however, this reference does not teach or suggest the refinement process of the present invention - that is, refining the sets of candidate prefixes and suffixes to identify actual prefixes and suffixes, including using knowledge, previously discovered in the refining step, to further refine said sets of candidate prefixes and suffixes. Instead, Wang, et al uses a statistical analysis to refine an initial language mode.

The other references of record have been reviewed, and these other references, whether considered individually or in combination, also do not disclose or suggest the refinement process of the present invention.

For example, Kanno describes a process for identifying a suffix. The Examiner cited this reference for the disclosure of reversing a word and adding it to a Patricia tree for suffix matching. There is no teaching, though, in Kanno of the above-described refinement process used in the present invention.

Independent Claims 1, 13 and 21 are being amended to describe this refinement process. In particular, each of these claims describes the feature that the affix determining process includes obtaining a collection of words, adding the words into a prefix Patricia tree, using said prefix tree to identify a set of candidate prefixes, reversing each of the words, adding the reversed words into a suffix Patricia tree, and using the suffix tree to identify a set of candidate suffixes. Each of Claims 1, 13 and 21 also include the limitation of refining the sets of candidate prefixes and suffixes to identify actual prefixes and suffixes, including using knowledge, previously discovered in the refining step, to further refine said sets of candidate prefixes and suffixes.

This feature of the present invention is of significant utility because it improves the candidate prefixes and suffixes through iterative refinement.

Because of the above-discussed differences between Claims 1, 13 and 21 and the prior art, and because of the advantages associated with those differences, Claims 1, 13 and 21 patentably distinguish over the prior art and are allowable. Claims 2-12 are dependent from Claim 1 and are allowable therewith. Likewise, Claims 14, 15, 19, 20 and 29 are dependent from Claim 13 and are allowable therewith; and Claims 22, 23, 27 and 28 are dependent from, and are allowable with, Claim 21.

The Examiner is, accordingly, respectfully asked to reconsider and to withdraw the rejection of Claims 1-11, 13-15 and 21-23 under 35 U.S.C. 102 and the rejections of Claims 12, 19, 20, 27 and 28 under 35 U.S.C. 103, and to allow Claims 1-15, 19-23 and 27-29. If the Examiner believes that a telephone conference with Applicants' Attorneys would be advantageous to the disposition of this case, the Examiner is asked to telephone the undersigned.

Respectfully Submitted,

John S. Sensny
John S. Sensny
Registration No. 28,757
Attorney for Applicant

Scully, Scott, Murphy & Presser, P.C.
400 Garden City Plaza – Suite 300
Garden City, New York 11530
(516) 742-4343

JSS:jy